DOCUMENT RESUME

ED 041 944 TM 000 036

AUTHOR	Hillerich, Robert L.
TITLE	Evaluation of Written Language.
REPORT NO	AERA-3-70
PUB DATE	Mar 70
NOTE	11p; Paper presented at the annual meeting of the American Educational Research Association, Minneapolis, Minn., March 1970
EDRS PRICE	EDRS Price MF-$0.25 HC-$0.65
DESCRIPTORS	*Composition Skills (Literary), *Effective Teaching, *English Curriculum, *Evaluation Methods, Experimental Teaching, Program Evaluation, Standardized Tests, Teaching Methods, Teaching Techniques, *Writing Skills, Written Language
IDENTIFIERS	*Glenview Illinois Public Schools, Spelling Writing Program

ABSTRACT

An evaluation procedure was formulated to ascertain the effectiveness of an emphasis on the clarity and interest appeal of a composition as opposed to its mechanical correctness in improving a child's written expression. A random sample of themes were submitted to a general evaluation of content by six criteria and a linguistic analysis by nine criteria was performed to evaluate mechanical correctness. Later, a standardized test consisting of a timed written response to a drawing was instituted for grades 1-8. Results from this evaluation procedure suggest that emphasis on many writing experiences, evaluated in terms of the clarity and interest of the content rather than the mechanics, should continue and that the standardized test is a reliable instrument for such evaluation. (PR)

ED041944

AERA 3-70

U.S. DEPARTMENT OF HEALTH, EDUCATION & WELFARE
OFFICE OF EDUCATION
THIS DOCUMENT HAS BEEN REPRODUCED EXACTLY AS RECEIVED FROM THE PERSON OR ORGANIZATION ORIGINATING IT. POINTS OF VIEW OR OPINIONS STATED DO NOT NECESSARILY REPRESENT OFFICIAL OFFICE OF EDUCATION POSITION OR POLICY.

Evaluation of Written Language

Robert L. Hillerich

Efforts in the teaching of English in the Glenview Schools for the past several years have been directed toward improving the written language of children through emphasis on many experiences in writing and focus on the clarity and interest-appeal of that writing as opposed to its mechanical correctness. Formal beginning of this approach is marked by the Spelling/Writing Program, tested at Rugen School in 1964-65 (Hillerich, 1968). While research evidence is lacking, professional thinking has been moving strongly in the direction of this emphasis (Nebraska, 1965; Tiedt & Tiedt, 1967; Braddock, 1963).

More recently, efforts in the Glenview Schools have been devoted to the development of means for evaluating progress toward this goal of improved written expression. Such evaluation is not included in standardized tests of achievement, and yet it is essential, not only to determine whether or not progress is being made toward improving skill, but also to indicate, by the mere fact of testing, that this is an activity of major importance in the school program.

TM 000 036

This paper presents the results of two years' efforts to evaluate progress in improving written language.

Procedure: 1967-68

Thousands of themes, from all grade levels in the Glenview Schools, remained from the collections made for the 1965-66 word count in spelling. A random sample of one hundred of these papers was taken from the fourth grade collection and another one hundred from fifth grade. The same kind of random selection was made from themes collected during 1967-68. The Table of Random Numbers was used in selecting papers from both years. Since only the first pages remained from the early collection, only first pages were used from the 1967-68 papers. All identifying data, such as names and dates, were removed, and themes from both years were Xeroxed so the age of the paper would not serve as a clue to the year it was written. Xeroxed copies were than randomized within grade level, making a total of two hundred randomly shuffled papers at grade four and another two hundred at grade five.

A team of eight teachers, grades 4-8, was provided released time to evaluate the themes in terms of established criteria.

Criteria: General Evaluation. Because of known difficulties in the subjective evaluation of creative writing (Braddock, 1963), a general evaluation of the content of each theme was made by two independent teams of two members each. These teams used the following six items and evaluated for each item on a scale of one to five points:

1. Unity of thought
2. Logical order of development
3. Smooth transition within and between paragraphs

4. Variety in vocabulary

5. Variety in sentence structure

6. Vividness and appropriateness of expression

After each team had established a total grade (the sum of the six items, ranging from scores of 6 to 30) for each paper, they were forced to "normalize" their grades by changing the highest 10% (20 papers) to the perfect score of 30, the next highest 20% to a grade of 24, 40% to 18, 20% to 12, and the poorest 10% to a score of 6. The final score for each paper was an average of the "normalized" scores of the two independent teams.

Criteria: Objective Evaluation. Mechanics were evaluated by the teams as specific counts of the number of spelling errors and the number of capitalization punctuation errors.

Linguistic analysis included (1) number of words, (2) number of T-units (Hunt, 1965), (3) average T-unit length[1], (4) sentence paterns by type, (5) nominals other than pronouns used as subjects (Loban, 1963), (6) number of subordinate clauses, (7) average clause length (Hunt, 1965), (8) ratio of subordinate clauses to all clauses (Hunt, 1965), and (9) number of figures of speech.

Results

Table 1 shows the results of these analyses of approximately 200 papers at each grade level, reported as mean scores by group. Several papers had to be removed, without knowledge of their date, at the general-evaluation stage because they could not be evaluated: lists of words, picture stories, etc.

As shown in Table 1, ability of the groups was comparable, with some

[1]The "T-unit" has been justified and explained by Hunt. Briefly, the "T-unit" substitutes for "sentence" but avoids problems generated by compound sentences, run-on sentences, and so on.

TABLE 1

Results of Analysis of Children's Writing:

1967-68 (Experimental) and 1965-66 (Control)

	Grade 4		Grade 5	
	Experimental	Control	Experimental	Control
Number of Subjects	95	99	100	99
Median IQ	116.	116.	112.	117.
General Evaluation	18.92	16.85	18.54	17.42
Spelling errors/100 words	2.63	3.76	2.94	4.36
Mechanical errors/100 words	2.55	3.23	4.28	4.46
Number of Words	70.16	60.41	74.76	63.18
Number of T-units	6.63	6.96	8.07	6.97
Average words/T-units	10.58	8.68	9.26	9.07
Number of Subordinate clauses	1.71	1.57	1.34	1.37
Ratio: Subordinate clauses/All clauses	25.71	22.79	16.60	19.71
Mean clause length	8.42	7.07	7.94	5.57

disadvantage to the fifth-grade experimental group. Nevertheless, the average general evaluation for grade four and grade five experimental groups was higher than the average for the control groups. This suggests that the quality of children's writing improved in the two years, despite the fact that the 1965-66 group had received some benefit from the current emphasis on writing.

The most interesting findings relate to the mechanics. Since 1965, the teaching of spelling has been changed to a word-list approach and the amount of time devoted to spelling has been reduced to three periods a week. Despite these changes--or <u>because</u> of them--the number of spelling errors is

lower for both experimental groups.

Likewise, the mechanical errors of capitalization and punctuation have been reduced, despite a definite shift in the English program which greatly deemphasizes the marking of children's papers for mechanical errors.

The number of words and number of T-units are not significant in themselves, since all papers were arbitrarily cut to the first page. However, the number of words per T-unit increased in the experimental group at fourth grade and showed no appreciable difference at fifth. As reported by Hunt, greater length of T-units is a measure of sophistication in written language. He reported an average T-unit length for grade four as 8.6 words; for grade eight, 11.5; and for grade twelve, 14.4 (Hunt, 1965, p.22). It can be seen in Table 1 that all groups exceeded Hunt's figures.

The Analysis of sentence patterns could not be interpreted and was dropped from the evaluation. The experimental group at grade four used twice as many of the simplest pattern (Noun-Verb) and of the most complex (Noun-Verb-Noun-Noun) as did the control group.

Only two nominals other than pronouns were used in all themes, one in the grade four experimental group and one in the grade five control group. This item, and figures of speech, were also dropped from the evaluation. The use of figures of speech seemed more a result of accident due to the type of writing rather than a measure of sophistication in control of language.

The number of subordinate clauses was greater for the fourth grade experimental group but approximately the same for the two grade-five groups. While the ratio of subordinate clauses to all clauses was greater for the fourth grade experimental group, it was less for that group at grade five. This latter item is also considered a measure of sophistication by Hunt, who found ratios of 22.2 at grade four and 28.8 at grade eight. Table 1

shows that both fifth-grade groups fall short of Hunt's averages while the experimental group at grade four exceeds Hunt's group.

Finally, mean length of all clauses is greater for the experimental groups than for control groups, although the difference at grade five is slight. Hunt found mean lengths of 6.6 at grade four, 8.1 at grade eight, and 8.6 at grade twelve. All Glenview groups except the fifth-grade control exceed Hunt's averages, with the grade-four experimental group surpassing Hunt's eighth-grade average.

Procedure: 1968-69

Lack of controls, in terms of topic and time allotment, was suggested as a weakness of the 1967-68 comparison; therefore, in May, 1968, a standardized test of writing was instituted for all grades, 1-8. This consisted of a line drawing for each grade, along with standardized directions and time limits. This 1968 testing, although already influenced by the existing emphasis on written language, became the normative base for 1969 and future evaluations.

Primary Evaluation. Fluency is considered one criterion of skill in written expression. Hence, at the primary level, evaluation was based entirely on the number of words written in response to a test picture within a fixed time limit. Limits were set at thirty minutes for first grade, and twenty minutes for second and third. Norms were established from results of testing in May, 1968, using the total primary population (N= 1332).

Testing in 1969 followed the same procedure as in 1968. One hundred themes from each grade level, 1-3, were randomly selected and the number of words counted. Results are shown in Table 2.

TABLE 2
Results of Fluency Count in Primary Grades (1969)

Grade	Mean Words	Percentile (1968 Norms)
1	46	57
2	90	66
3	128	74

These findings indicate very clearly that primary children showed an increase in fluency of writing between 1968 and 1969. Observation during the school year suggested that teachers were having children do more writing than they had in the past. Since the testing done in 1968 was the first such testing of fluency at the primary level, it may have encouraged this increase in writing.

Upper Grade Evaluation. For grades 4-8, a random sample of 100 test themes per grade level was selected from the test writings of 1968 (control) and 1969 (experimental) classes. These were, in turn, randomized into a collection of 200 themes per grade, so that evaluators would not know the year in which they were written. Two teams of ten teachers each were provided released time to evaluate the test themes in terms of the criteria established and used in the 1968 evaluation.

Results. Table 3 shows the results of these analyses of approximately 200 papers at each grade level, reported as mean scores by group. Several papers had to be removed, without knowledge of their date, at the general-evaluation stage because they did not fit the scoring criteria (lists of words, poems, etc.).

As shown in Table 3, ability of most groups was comparable, with some disadvantage to the fourth and fifth-grade experimental groups. There was no appreciable difference among the groups in the general evaluation, all

TABLE 3

WRITING EVALUTION -- May, 1969

Random Samples: 1967-68 (Control) & 1968-69 (Experimental)

	Grade 4		Grade 5		Grade 6		Grade 7		Grade 8	
	Exp. N=100	Cont. N=99	Exp. N=100	Cont. N=100	Exp. N=99	Cont. N=99	Exp. N=97	Cont. N=99	Exp. N=100	Cont. N=100
Intelligence	111	116	112	114	114	114	114	114	114	115
General Evaluation	18.00	18.12	18.12	17.88	17.76	18.06	17.94	18.00	18.00	18.12
Number of Words/theme	163.35	154.10	176.73	188.58	217.70	216.69	210.35	199.77	191.47	171.15
Spelling Errors/100 words	4.37	4.27	3.80	3.43	3.20	2.33	2.19	2.08	1.51	2.06
Mech. Errors/100 words	5.98	4.63	6.63	6.17	2.95	2.97	4.02	3.98	3.16	3.32
Number T-units/theme	16.20	14.94	16.93	17.96	19.82	19.85	18.32	16.26	16.44	15.24
Average Words/T-unit	10.08	10.32	10.44	10.50	10.98	10.92	11.48	12.28	11.65	11.23
Number of sub. clauses	1.85	1.63	1.21	1.32	5.70	4.79	2.96	2.88	4.09	3.90
Number words in sub. cls.	11.13	10.02	7.00	8.59	38.21	33.12	17.33	17.61	29.53	27.85
Average/clause	6.02	6.16	5.79	6.50	6.71	6.91	5.86	6.12	7.22	7.14

ranging less than 0.2 points from the average of 18.0.

In terms of fluency, there was no significant difference in number of words written by the groups at sixth grade. At all other levels, except fifth, experimental groups exceeded control. Number of words cannot be compared to the 1968 evaluation, since that year included only the first page from each theme.

Spelling errors were about the same for experimental and control groups at all grades except sixth and eighth. The experimental group at grade six had slightly more errors and at grade eight, slightly fewer. Spelling errors tended to decrease with grade level. In comparison to the previous evaluation (Table 1), fourth grade showed more spelling errors and fifth grade showed about the same number (more than the experimental and fewer than control groups in the 1968 evaluation). This comparison is not a true one, however, because the 1968 evaluation included only the first page from each theme. A check of fourth grade themes, in 1969, revealed that the second page of those themes averaged one spelling error and one mechanical error more, per 100 words, than the first page of the same paper. Hence, the longer the theme of a given child, the more errors per hundred words.

No difference existed between groups in terms of mechanical errors except at grade four, where the experimental group averaged 1.3 more errors.

No appreciable differences existed between experimental and control groups in terms of length of T-units, but a slight difference in terms of number of T-units favored the experimental groups. This fact is a reflection of the greater number of words written by the experimental groups. In comparison with the 1968 evaluation, length of T-units increased at grade 5, and, at both grades 4 and 8, exceeded the length reported by Hunt (1965)

for those grades.

In terms of subordination, the experimental group in grade six used significantly more subordinate clauses and therefore had more words in subordinate clauses; however, average clause length was about the same for experimental and control groups at all levels and tended to increase slightly at each grade level.

Discussion

The first evaluation of written language (May, 1968) indicated considerable improvement over 1966, despite the fact that the spelling/writing program had already affected the control group.

The 1969 evaluation revealed dramatic progress at the primary level, where writing is measured only in terms of fluency. Undoubtedly this is a result of the reported increase in the amount of writing done by children during the year.

While evaluation at grade 4-8 revealed a few more advantages to the 1969 group, the results were not so dramatic. This should not mean that we have reached a saturation point with children. It could mean, as has been observed in the classrooms, that middle-grade children were not writing as much. During 1967-1968, when the English program did not include transparencies and pupil pages, there was more writing done by children than was true in 1968-1969.

In conclusion, efforts to evaluate written language in the Glenview Schools have produced results in two areas. First, the evaluations themselves suggest that progress has been made since 1965-66 in terms of increased skill and sophistication in written language. Hence, in this respect we conclude that emphasis on <u>many</u> writing experiences, evaluated in terms of the clarity and interest of content--as opposed to emphasis on

mechanics and the use of red pencils--should continue.

Secondly, the major concern reported in the 1968 evaluation has been resolved: we have a standardized test for written language and we have two years of experience with it. This test serves in several ways: (1) it enables annual evaluation of progress in the written language program; (2) it provides an opportunity for in-service activity on the part of teachers doing the evaluating; and (3) knowledge that it is scheduled annually serves to encourage more experience in written language in the classrooms.

REFERENCES

1. Braddock, Richard et al, Research in Written Composition, National Council of Teachers of English, 1963

2. Hillerich, Robert L., "A Spelling/Writing Project Becomes an English Program," Illinois ASCD Newsletter of Curriculum and Supervision, March, 1968, pages 1-3.

3. Hunt, Kellogg W., Grammatical Structures Written At Three Grade Levels, National Council of Teachers of English, 1965.

4. Loban, Walter D., The Language of Elementary School Children, National Council of Teachers of English, 1963.

5. Nebraska Curriculum Development Center, University of Nebraska, 1965.

6. Tiedt, Iris M. and Sidney W., Contemporary English in the Elementary School, Prentice Hall, 1967.